A 6DoF Fiducial Tracking Method Based on Topological Region Adjacency and Angle Information for Tangible Interaction

Hiroki Nishino
Graduate School for Integrative Sciences & Engineering, National University of Singapore
CeLS, #05-01, 28 Med. Drive, Singapore 117456
g0901876@nus.edu.sg

ABSTRACT
In this paper, we describe a new method for camera-based fiducial tracking. Our new method is based on the combination of topological region adjacency and angle information, where as related works by Johnston's *RAG target* [7], Costanza's *D-Touch* [3], and Kaltenbrunner's *reacTIVision* [2] are based on the uniqueness of the topological region adjacency structure.

Such a combination of the topological region adjacency and angle information enables a wider unique ID range, while maintaining the merit of fast and robust fiducial tracking in topology-based approach. Our method makes it possible to obtain the 6 degrees-of-freedom (6DoF). Such problems of a narrow unique ID range and lack of 6DoF information have been the main deficits in most systems based on topological region adjacency approach, when compared to other fiducial tracking methods.

Author Keywords
Fiducial tracking, computer vision, tangible interaction, human computer interface.

ACM Classification Keywords
H.5.2. User Interfaces, I.4.9. Image Processing and Computer Vision Applications

General Terms
Experimentation, Algorithms, Design

INTRODUCTION
Camera-based fiducial tracking has been widely used as a method to implement tangible interaction systems. While markerless object detection techniques that utilize SIFT [9] and its family are rapidly gaining popularity among augmented reality community, marker-based fiducial tracking techniques still maintain a considerable importance for tangible interaction. Marker-based fiducial tracking can achieve considerably faster processing speed and easily identify a large number of the objects with their own unique IDs than currently available markerless tracking techniques. Such features of marker-based tracking are quite important for tangible interactive systems, e.g. to detect quick user gestures that are applied to many tangible objects at once.

Among many fiducial tracking systems, *reacTIVision* and *D-Touch* have been rapidly gaining popularity among tangible interaction applications. Both of them are based on the topological region adjacency approach.

The topological region adjacency approach can achieve fast processing speed and the robustness against false positive detection easily, yet usually existing systems based on this approach still have the deficit of a narrow unique ID range and cannot provide 6DoF information, when compared to those based on matrix-pattern or pattern-matching; A narrow ID range can be an obstacle for collaborative interaction especially when many personalized objects owned by the users need to be distinguished uniquely. 6DoF information can be useful to detect more complicated gestures applied to the interface objects in a workspace.

To improve the topology-based approach for better use in tangible interaction, we developed a novel method that combines topological region adjacency and angle information between regions inside a marker. Our new method can extend the unique ID range much wider and enable 6DoF pose estimation, while maintaining the merits of the fast processing speed and the robustness of a topology-based approach.

RELATED WORKS

Matrix-Pattern and Pattern-Matching Approach

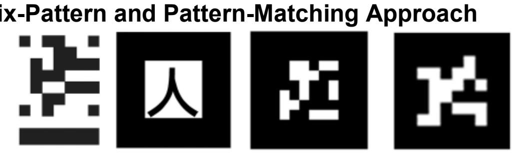

Figure 1. The examples of the markers from *CyberCode* [1], *ARToolkit*[8], *ARToolkit Plus* [12] and *ARTag* [6] (from Lto R)

The use of matrix-pattern to encode IDs can be frequently seen in fiducial tracking systems. There is already a substantial body of previous research, such as *CyberCode* [10], *ARToolkit Plus* [11] and *ARTag* [5]. Also there are other fiducial tracking systems that use pattern-matching to

TEI 2010, January 25–27, 2010, Cambridge, Massachusetts, USA.

detect a fiducial, such as *ARToolkit* [8]. All of these can obtain 6DoF information. The examples of these markers are shown in Figure 1 in the previous page.

Generally speaking, those based on a matrix-pattern usually encode one binary digit into each cell and have a larger data capacity than those based on a topology-based approach. Pattern-matching methods have the merit of visually meaningful marker designs. However, such types of systems usually require some digital techniques to improve the robustness to avoid false detection as described in the later section.

Topological Region Adjacency Tree Approach

The use of topological region adjacency for fiducial tracking has been seen in the previous research such as Johnston's *RAG target* [7], Costanza's *D-Touch*[3] and Kaltenbrunner's *reacTIVision* [2]. Among these, only *Rag target* can provide 6DoF information. Figure 2 shows the examples of the fiducial markers from these systems.

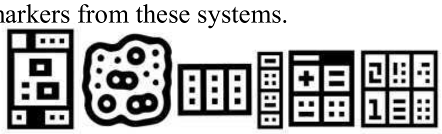

Figure 2. The examples of fiducials in RAG target [7] (left), reacTIVision [2] (2nd from left), and D-Touch [4] (right 4)

The basic idea in topology-based approach is to express the containership information of the regions in black/white image as a graph of region adjacency. *D-Touch* and *reacTIVision* express such region adjacency as a tree. Figure 3 is an example taken from Kaltenbrunner's paper [2]. The containership information of the fiducial on the left can be expressed as a tree on the right.

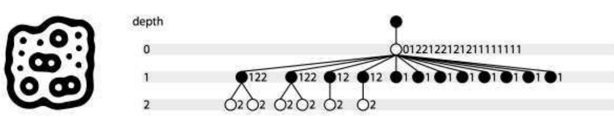

Figure 3. A reacTIVision marker and its topological region adjacency tree with left heavy depth sequence [2].

In *reacTIVision*, such a tree of topological region adjacency is interpreted into a canonical sequence called *left heavy depth sequence* and used to map a tree structure to an ID of a fiducial. Such a canonical sequence is necessary since region adjacency tree does not supply anything more than the containership information. In Figure 3, the tree on the right can be as *left heavy depth sequence* of 0122122121211111111 and map to its own ID.

Generally speaking, in such an approach based only on the topological region adjacency tree, the range of unique IDs is essentially limited by the possible combination of sub-trees with different topological features. This is one of the reasons why topology-based fiducial tracking systems have less unique IDs than matrix-based systems.

The Techniques to Increase Robustness

The trade-off relationship between the complexity of a fiducial marker design and the robustness against false detection is a significant issue. This is quite apparent in matrix-pattern methods. For instance, a white cell in a marker might disappear and be misinterpreted as black when the quality of input is not sufficiently good. Not to let such misinterpretation result in false detection, matrix-based tracking systems usually involve digital techniques such as CRC, hamming distance and the like, as seen in [5].

A topology-based approach usually does not require such techniques for robustness against false positives, if the given topological structures are sufficiently complex and hardly exist in the input image except fiducial markers. Such a study in the complexity of topological structure and the robustness is well described in [3]. However, it is still important to avoid using fiducials that have a topological structure alike to others, to improve robustness against inter-marker confusion.

DESCRIPTION OF OUR METHOD

We describe our new method for 6-DOF fiducial tracking in this section. Our method is based on the combination of the topological region adjacency and the angle information. Since our method is based on the topological region adjacency of black/white image, it is required to binarize and segmentize image input to obtain toplological structures. Figure 4 shows the pictures from each phase in our system.

Figure 4. images form each phase. input image, gray scaling, binarization and detection (from left to right)

Fiducial Recognition

Figure 5 shows three examples of our 16bit fiducial markers. Our fiducial consists of two main components.

The first one is the black square in the center, which contains one white dot. This is used to obtain the rough angle of the fiducial in the video input image, using the vector from center of the minimum bounding box of the black square to the white circle. The black and white regions are surrounding this center black square. Each dot in those regions encodes a bit, 0 for a black dot and 1 for a white dot. These bits are sorted in a clockwise order, starting from the rough angle obtained from the center circle, to decode the ID of the fiducial. Decoding these examples in Figure 5 results in their unique IDs, 48115, 64407, 40879, from left to right. Notice all these fiducials has the same topological structure. In case of 16bit fiducial, there are only 17 different topological structures, since the number of black or white bit-encoding nodes can vary only between 0-16.

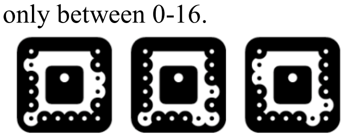

Figure 5. Three example of our fiducials

We seek for such a topological structure that contains the 16 bit-encoding nodes and the center square to find a fiducial candidate and check if the angles betweens the bit-encoding dots are valid. All the required information can be obtained during segmentation. The 4 corner bit-encoding dots can be used for 6DoF pose estimation as in Figure 6.

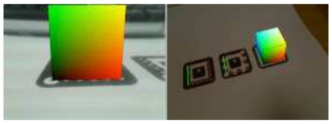

Figure 6. Overlaying a 3D cube using 6DoF information

Robustness of Our Method

As described in the previous section, topological region adjacency approach takes very different strategy to increase the robustness when compared to matrix-based fiducial tracking systems. While those systems that encode data as matrix pattern depend their robustness largely on digital techniques such as CRC or hamming distance, those topology-based systems depend on the rarity of the topological structures in input images. As in the case of *reacTIVision* [2], a topological structure with the depth of 3 and 19 tree nodes is sufficiently rare to avoid false positive detection.

In our fiducial design shown in Figure 5, the topological region adjacency tree of our fiducial has the depth of 4 and 20 nodes. Combined with additional validation with angle information, our method can be significantly robust against false positive detections. The robustness against inter-marker confusion in our method is supported by having enough margin space around the bit-encoding nodes. As described in the next section, our prototype is robust enough for most of practical applications of tangible interaction.

EVALUATION

We evaluated the current prototype system in our lab environment[1] under a normal lighting condition. We also compared our prototype to *reacTIVision* [2] in several experiments, since it is one of the most widely used fiducial tracking systems in tangible interaction community.

Unique ID Range

Our prototype is currently implemented with 16bit data capacity, which is considerably larger than other topology-based systems. The number of unique fiducials in *RAG target* is 449 [7], and *reacTIVision* has 180 unique fiducials [2]. In *D-Touch*, there are several different types of fiducials as in Figure 2; each of them has 1 unique ID, 6 unique IDs, 24 unique IDs, and 120 unique IDs [4]. Such a narrow ID range has been one of the deficits of topology-based system; especially when compared to the number of unique IDs of matrix-based systems (*ARTag* 2002; *ARToolkit Plus* 4096; *Cybercode* 24bits)

Fiducial Recognition Speed

We measured the average processing time cost of 1000 frames with 12 markers in input images with *capture (a) in 640x480* resolution. *ReacTIVision* also has several features to increase the robustness like frame equalizer, the use of information from the previous frame and so on. Taking these features into the evaluation can increase the time cost. To compare these two different systems, we excluded the time cost for *reacTIVision* features that our system has not implemented yet. Table 1 shows the result.

	binarization	**segment-ation**	**fiducial detection**	**total dur.**
our system	1.764	1.937	0.058	3.759
reacTIVision	1.632	1.682	0.076	3.390

Table 1. The average processing time per frame in msec (the values are rounded off to 3 decimal places)

The only significant difference in the algorithm between our prototype and *reacTIVision* is the phase of fiducial detection, and both apply the same algorithm to the other phases such as binarization and segmentation, but in different implementations. Taking such implementation/optimization issue into account, our method doesn't have any significant damage to its performance compared to *reacTIVision*, while it can provide a much wider ID range and 6DoF information.

False Positive Detection

We evaluated the robustness against false positives of our prototype system and *reacTIVision*, with four different video inputs[2], all of which do not contain any fiducial. Table 2 below shows the result. Our method is significantly robust against false positives and showed a better result than *reacTIVision*. In *film*, *reacTIVision* detected most of the false positives in the very last part of its end credits. For reference, we also put the number that excludes these false positives in the last part inside parentheses.

	room	**outdoor**	**film**	**anime**
our system	0	0	0	0
reacTIVision	2	0	129(31)	21

Table 2. The number of the false positives observed

Inter-marker Confusion

In our method, when any disappearance of a region causes a change of topological structure, it prevents a marker to be recognized as a candidate because its topological structure is no longer one of 17 valid structures. Yet inter-marker confusion can occur if the given margin space around a bit encoding-dot is not sufficiently large. In Figure 7, the result

[1] Dell Desktop PC : Intel Core 2 Quad/2.5GHz, 4GB MEM, Windows Vista Home Premium, Visual C++ 2008 Express
Capture(a): Logicool QCam Pro WebCam
Capture(b):Alphawireless EZCAP with NTSC video input
Capture(c):Alphawireless EZCAP with PAL video input

[2] room : 5 minutes video, taken inside our lab - *Capture(c)*
outdoor : 5 minutes video, taken outdoor, - *Capture(c)*
film : the 136 minutes of the film, *Matrix* - *Capture(b)*
anime : the 25 minutes of the animation, *Ghost in the Shell, Stand Alone Complex: Episode No.4* - *Capture(b)*

of the binarization should be as (a), but if the black bit-encoding dot were absorbed to the other area because of some noise in input, creating a white dot as in (b); this causes a swap of the color that results in detection of a wrong fiducial ID.

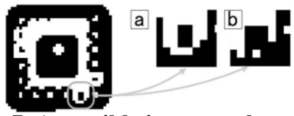

Figure 7. A possible inter-marker confusion

To avoid such false detection, we place the bit-encoding dots in the circular margin space, the diameter of which is 3 times as large as the bit-encoding dots. This margin space is likely to be enough to cause no inter-marker confusion in usual webcam input, yet it is still necessary to consider input with undesirable noise that results in false detection.

We ran 2 experiments related to such false detection. For *test 1*, we took 1000 frames of 640x480 pixels with 3 fiducials on the desk with ID=0, 65535(all 1), 43690(1010101010101010) by *capture(a)*. The camera was moved several times between the closest and the largest distance in which at least one of these fiducials can be detected. Then we added noise to see its robustness against given noise. In *test 2*, we used 1000 frames, freely moving the camera in various angles and distances to simulate a practical augmented reality application. We generated random numbers of normal distribution (the average=0 and variances=0.2) and uniform distribution (between -1 and 1), then converted them to 0-255 range and mixed them into 8bit gray-scaled input images in several different ratios.

In Table 3, *noise level(%)* is the ration that added noise occupies in the mixed result images. The values outside parentheses are the number of inter-marker confusion and those inside parentheses are the number of the correct detection. There was no inter-marker confusion observed in the original images in both tests. The total number of detected fiducial in the original image is 2408 in *test 1* and 2639 for *test 2*.

noise (%)

variance

	1%	**3%**	**5%**	**10%**
Test 1				
normal	0(2412)	2(2336)	1(2208)	17(1615)
uniform	3(2370)	6(2058)	10(1623)	3(919)
Test 2				
normal	0(2655)	1(2659)	3(2581)	11(2079)
uniform	0(2641)	3(2363)	10(1892)	16(805)

Table 3. The number of inter-maker confusion

The addition of noise seems to mostly result in less marker detection, yet there is occasional inter-marker confusion observed for very noisy input. Applying digital techniques such as CRC or the use of previous frame information may be desirable for noisy input.

CONCLUSION

We developed a novel camera-based tracking method, by combining the topological region adjacency and the angle information. It provides a wider ID range and 6DoF information, while maintaining the fast processing speed and the robustness of the topology-based approach. These features are beneficial for tangible interactive applications. The lack of such features has been one of the major deficits of existing topology-based systems.

FUTURE WORKS

We are currently re-designing the marker shape to improve its spatial cost. The current design requires about 40x40 pixels at least for detection, but a circular fiducial design seems to reduce the minimum size to around 30x30 pixels.

ACKNOWLEDGEMENT

The authors thank Information-Technology Promotion Agency Japan for supporting part of this software development by Exploratory Software Project Funding.

REFERENCES

1. Ayatsuka,Y. and Rekimoto, J. (2006). Active CyberCode : a directly controllable 2D code. *Proc of CHI'06 extended abstracts on human factors in computing systems*
2. Bencina, R., Kaltenbrunner, M. and Jorda, S. (2005). Improved Topological Fiducial Tracking in the reacTIVision System. *Proc of IEEE Procams'05*
3. Costanza, E. and Robinson, J. (2003). A region adjacency tree approach to the detection and design of fiducials, *Proc of VVG 2003,* pp.63-69
4. Costanza, E. Enrico Costanza – Research http://web.media.mit.edu/~enrico/research/research.php ?projectTitle=Older%20Markers, accessed on Apr/24/09
5. Fiala, M. (2005) ARTag, a Fiducial Marker System Using Digital Techniques, *Proc of CVPR'05*, pp590-596.
6. Fiala, M., ARTAG http://www.artag.net/, accessed on Apr/24/09
7. Johnston, D. (2001). Untethered VR. Ph.D thesis, University of Essex
8. Kato, H. and Billinghurst, M. (1999). Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System. *Proc. Of ISMAR99*
9. Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Key Points, *Int'l J. Computer Vision, vol 60. no.2 pp.91-110*
10. Rekimoto, J. and Ayatsuka, Y. (2000). CyberCode: Desigining Augmented Reality Envrionments with Visual Tags. *Proc of DARE'00*
11. Wagner, D. and Schmalstieg, D. (2007). ARToolkit Plus for Pose Tracking on Mobile Devices. *Proc of CVWW'07*
12. Wagner, D. ARToolkit Plus, http://studierstube.icg.tu-graz.ac.at/handheld_ar/artoolkitplus.php, accessed on Apr/24/09